**DATASET GUIDANCE**

**Promotion of open data**

It is widely accepted that the publication of public data can provide a local economic stimulus. There is a high demand for council data from developers who wish to create apps and websites to provide new and innovative ways to present the data and provide services. Analysts are also keen on working with the data to provide new insight into how the city functions and potential solutions to some of its problems. It may also be useful to ordinary members of the public who simply want to view the data.

Leeds City Council, where possible, is publishing its datasets using an Open Government Licence. This allows developers to re-use the data in any way they wish, subject to a few conditions, and protects the council at the same time. Other similar licences are available (e.g. Creative Commons).

**Identification of datasets**

This could come from a variety of sources. It may be that an organisation wants to simply provide public access to their data in respect of being open and transparent or that national legislation compels them to publish a particular dataset.

The Information Asset Owner will need to be consulted to discuss the publication of *their* data. They are ultimately responsible for the data collected and used by their service and will also be responsible for its publication. The integrity and accuracy of the data will also need to be considered. It is important that what is published is usable, understandable and accurate.

Information management and governance officers need to work with colleagues to review accuracy and quality and to ensure that no personal and/or business sensitive data is published and that data matching is not possible.

**Processing the data prior to publication**

The processing of anonymised data falls outside of the Data Protection Act 2018 and General Data Protection Regulation (GDPR) except in relation to the ‘right to be informed’ which requires you to include the relevant information in your privacy notice.

For those organisations who need to respond to Freedom of Information requests, changes to the Freedom of information Act in September 2013 now states that if a request is made for a dataset (they don’t need to specifically use that phrase), then this should not only be made available to the requester, but also published and maintained on an organisation’s website and/or open data repository (Data Mill North).

**Dataset rules**

You cannot publish datasets on the Data Mill if the information is not anonymised (refer to the anonymisation guidance below).

Data should be published in an open and re-usable format. Leeds City Council has informally adopted CSV (Comma Separated Values) as this standard, though others are in common use. CSV can be saved and opened up as a spreadsheet, but means that those people who do not purchase proprietary products can still access the data.

It is important that sufficient information is provided to provide context to the data being published. This ‘data about the data’ is referred to as metadata. Where possible, organisations should provide as much detail as possible when uploading their data to Data Mill North.

* Columns: headers should not be duplicated; column headings should always remain the same – i.e. do not change them on later versions; always add new columns at the end, do not ‘slot in’ new columns in later versions;
* Never publish datasets with any hidden columns/rows or pivot tables; always double check before publication;
* Use CSV format as detailed above;
* Always try and include some geographic information e.g. longitude & latitude, eastings & northings, post codes etc. and consider if there are any Royal Mail or Ordnance Survey licensing restrictions;
* Don’t change formats – i.e. if publishing in CSV, ensure this is always the case;

**Maintaining the datasets**

Once a dataset is published, it needs to be maintained. Failure to publish up to date and relevant information could result in complaints, queries or the return of Freedom of Information requests. Information Asset Owners will need to be made aware of their commitment to the continued publication of a dataset and that processes are in place to ensure the dataset is maintained and updated.

**Training**

The Open Data team will provide training to those individuals and organisations who will be publishing to Data Mill North.



**What is anonymisation?**Anonymisation is the removal of information that could lead to an individual being identified either directly, or indirectly in combination with other available information. Anonymisation is of particular relevance now, given the increased amount of information being made publicly available through Open Data initiatives and through individuals posting their own personal data online.

**How can a person be identified from data?**The General Data Protection Regulation (GDPR) and the Data Protection Act 2018 regulate how we process "personal data". In order to publish anonymised data on the Data Mill, any “direct or indirect identifiers” must be removed to mitigate the risk of breaching the above laws.

**Direct identifiers** include names, addresses, postcodes, telephone numbers or pictures; account reference numbers, e.g. council tax, NHS or National Insurance number; URLs and cookies.

**Indirect identifiers** which, when linked with other publicly available information sources, could identify someone, e.g. information on workplace, occupation or exceptional values of characteristics like salary or age.

**How can I be sure that I have completely anonymised information?**As a general rule if there are at least 6 individuals or more to whom the information could refer, then this could be classed as anonymous. However this will depend on what information/data you are publishing, e.g. if your data relates to an individual of a specific gender and ethnicity living at a certain postcode you can increase the number of people to whom it could refer by only using the first 3 digits of the postcode.

**Where can I find more information?**

[The information Commissioner’s “Anonymisation: managing data protection risk code of practice”](http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation)

[The UK Anonymisation Network (UKAN)](http://www.ukanon.net/)